DOCUMENT RESUME

ED 128 397 TM 005 552

AUTHOR Benton, Sidney E.; Scott, Owen
TITLE A Comparison of the Criterion Validity of Two Types of Student Response Inventories for Appraising Instruction.
PUB DATE [Apr 76]
NOTE 37p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (San Francisco, California, April 1976)

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS College Students; College Teachers; *Comparative Analysis; Criteria; Effective Teaching; Factor Structure; Grade Point Average; Higher Education; Prediction; *Rating Scales; *Statistical Analysis; *Student Evaluation of Teacher Performance; Test Construction; Test Reliability; Test Selection; *Test Validity
IDENTIFIERS *Inventory of Student Perceptions of Instruction; *Student Instructional Report

ABSTRACT

Two types of student response inventories for appraising instruction, the Student Instruction Report (SIR) and the Inventory of Student Perceptions of Instruction (ISPI) were administered to 554 students in 30 classes at two Georgia colleges. Mean responses to items were used in principal components analyses with varimax rotation. The two instruments were then administered to 525 students in 31 English classes having a common final examination. Results of multiple regression analyses produced a significant multiple correlation with one instrument and self-reported cumulative grade point averages, with end-of-course grades in English as the criterion variable. Results of this study suggest that student appraisal of college instruction using SIR in conjunction with self-reported grade point averages are of some help in predicting end-of-course final examination scores. If the effectiveness of an instructor is measured in terms of end-of-course achievement of his class, then college administrators should proceed with caution in using student ratings to gauge instructor effectiveness. Moreover, results of this study lend some support to the use of instruments developed empirically over those developed rationally. (Author/BW)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED128397

A Comparison of the Criterion Validity of Two Types of Student Response Inventories for Appraising Instruction

Sidney E. Benton

North Georgia College

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Owen Scott

University of Georgia

TM005 552

Paper presented on April 20, 1976, at the annual meeting of the National Council on Measurement in Education in San Francisco, California.

Many faculty members criticize the use of student rating forms. However, after a review of the empirical studies pertinent to these criticisms, Costin, Greenough, and Menges (1971) concluded that the ratings can provide reliable and valid information on the quality of courses and instruction. They further pointed out that "faculty resistance to the use of student rating forms may stem partially from the fact that many rating forms have been prepared by groups or individuals not qualified to construct such instruments" (p. 511). This claim seems founded. Miller (1972) stressed, "Too many procedures for evaluation consider only the first step, the development of evaluative criteria." (p. 15)

While a review of the literature revealed the widespread use of student evaluation instruments, little information about the genesis of the instruments was given. Further, researchers who have studied student ratings have generally been vague about the characteristics of the instruments they chose to study and the reasons for those choices. Since college instruction is an important aspect of our society and student rating of instruction is a common practice, research pertaining to the instruments used is very worthwhile.

Considering all of the checklists and forms available, it is understandable that it would be difficult for the average instructor to select one instrument that he could use and have confidence in using. A partial solution to this problem is to: (a) survey student evaluation instruments currently in use, (b) select the best of these instruments, and (c) check the criterion validity of the selected instruments.

The literature review identified two approaches for developing student rating instruments: a rational approach and an empirical approach. The rational approach is similar to the approach used in the construction of an achievement instrument with items written in terms of a set of specifications. On the other hand, in the empirical approach, items are collected from a variety of sources, and principal component analysis, or some type of factor analyses, is then used to categorize the items. Most researchers apparently stress the empirical approach, but there is no evidence to show which is the better approach or to what extent a combination of these two approaches is possible or advantageous. Whether one set of scores yields a better prediction of achievement than another should be of concern in instrument development. A dependable answer to this question might suggest future approaches to construction of instruments of students' assessments of college instruction.

OBJECTIVES OF THE INQUIRY

The overall purpose of the study was to compare two types of student response inventories for appraising college instruction. A major aspect of the comparison of the two instruments involved criterion validity checks. In the criterion validity checks, a major goal was to determine whether scores based on components or scores based on rational sections were better predictors of end-of-course achievement.

The specific purposes of this study were to:

1. Identify instruments currently being used for student evaluation of college instruction.
2. Develop a set of criteria to be used in selecting instruments for possible use in the study.

3. Select two instruments for use in the research--one instrument developed rationally and the other instrument developed empirically.
4. Administer the two instruments to an available pool of classes.
5. Analyze the responses to the two instruments by principal components analyses.
6. Readminister the two instruments to classes that had a common end-of-course examination.
7. Use self-reported grade point averages (GPAs) together with mean raw scores based on: (a) the rationally defined categories of the first instrument, (b) the empirically derived categories of the first instrument obtained by the principal components analysis, and (c) the categories of the second instrument obtained by the principal components analysis, to predict students' end-of-course achievement.
8. Report and summarize the comparisons of the two instruments.
9. Make recommendations (or suggestions) on the basis of these comparisons.

SELECTION OF INSTRUMENTS

A comprehensive search of the literature in the field was undertaken to identify instruments currently being used for student evaluation of college instruction. Sources included ERIC (1970-1973), Buros' Mental Measurements Yearbooks, Buros' Tests in Print, relevant papers presented at the American Educational Research Association and National Council on Measurement in Education annual meetings, and other appropriate bibliographic sources. Programs of these annual meetings were examined for all topics concerned with student appraisal of college instruction and requests for these papers were made by mail.

Thirty-nine instruments for appraising college instruction were located through the search of the literature. To be included in this list, an instrument had to be applicable to any academic area, designed to evaluate college teaching, and provide information to be used in the improvement of instruction. In order to reduce the number of instruments, a list of criteria to be used in evaluating the instruments was developed by the present writer. Feedback on these criteria was obtained from graduate students, professors, and students in a graduate seminar in research. The list was revised and the final list of criteria is as follows:

CRITERIA FOR FIRST SELECTION OF AN INSTRUMENT

1. The instrument was designed to evaluate any college level course in an academic classroom setting.
2. An implied or stated purpose of the instrument is to provide specific information to the instructor for use in improving teaching.
3. Data on the instrument are available in publications and/or from papers presented at national meetings of scholarly organizations.
4. The instrument is available for purchase or may be reproduced by obtaining permission from the author.
5. The instrument is one that is not unreasonable in terms of time and effort for administration.
6. Items in the instruments were either selected inductively using responses from groups comparable to one for which the instrument is intended or items were selected on the basis of clearly described rationale.
7. Reliability information has been reported.

8. Validity information, in addition to content validity, has been reported.

The properties of each of the thirty-nine instruments were then examined to determine which of the instruments met all of the criteria.

The Inventory of Student Perceptions of Instruction (ISPI) by Scott (1973) was the only instrument identified that has the content developed rationally rather than empirically and that at the same time satisfied all of the set of predetermined criteria. Most ISPI items are descriptive rather than evaluative and cover most significant aspects of instruction. When compared with other student evaluation of college instruction instruments, ISPI is as comprehensive as other available instruments.

From the instruments that met all criteria, two instruments with empirically developed content were selected. These instruments were those that were widely used and widely reported in the literature. Four judges then selected one of the two instruments on the basis of the list of criteria and on the basis of the instrument that they would choose to administer in their classes. An agreement of three of the four judges constituted the basis of selection of the Student Instructional Report (SIR) by Centra (1972) as the second instrument to be used in the study.

These two instruments (ISPI and SIR) selected for the study were the two, then, that best exemplified the rational and the empirical approaches to developing such instruments.

METHODS

In the winter quarter of 1974, the two instruments were administered to 554 students in 30 classes at the University of Georgia and Brenau College.

The instructors were told that the purpose for obtaining the student ratings in their classes was to analyze the responses to the two instruments using principal components. They were assured that the purpose was not to compare their teaching effectiveness with other instructors and that no one would know their identity. These instructors were told that they could obtain results of their ratings from the investigator.

Methods suggested by several writers—Scott (1973); Costin, Greenough, and Menges (1971), and Eble (1971)--were considered and were incorporated in the following list of procedures for administration of the two forms:

1. The forms were administered during the last week of classes.
2. The forms were randomly distributed with one half of each class receiving ISPI and the other half of each class receiving SIR.
3. Students responding to ISPI marked all 69 items and students responding to SIR marked 31 items (8 items were omitted from SIR since they are largely student descriptive information).
4. Students were assured that their identity would not be disclosed and that their participation would not affect their grades in any way. They were told that the instructors would not know the results until final grades had been posted.
5. Students were told that the ratings would give their instructors an opportunity to learn about the effectiveness of their teaching techniques and to make appropriate adjustments.
6. Each instructor was asked to leave the classroom while the senior author had the students fill out the ratings.

Mean responses to the items on the two instruments were then analyzed by principal components with varimax rotation. In advance of the principal components analyses, it was determined that items with primary loadings less than .40 and items loading high on more than one principal component (difference of the absolute values of the highest loadings had to be greater than .10) would be eliminated from the study. It was expected that the elimination of items with high loadings on more than one principal component would essentially identify principal components that were uncorrelated and thus yield higher multiple correlations in the prediction of final examination scores. Principal components obtained were used in the second phase of the study.

The following quarter, the two instruments were administered at the University of Georgia to 525 students in 31 freshman English classes that had a common end-of-quarter examination. Again a random half of each class was given one instrument and the other half, the second instrument. Purposes of the study were explained and the instructors were assured that if they participated in the study no one would know their identity. They were also told that results of the ratings would be made available to them.

In this administration, forms were attached to the answer sheets, and students were asked to supply their names, their identification numbers (IDs), their cumulative grade point averages (GPAs), and number of quarters they had attended the University of Georgia. They were told that the information on this form was for research purposes only and that in no way would they be identified. Other procedures were similar to those of the previous quarter with the exception that data were collected during the last two weeks of the quarter rather during the last week.

Rayder (1968) reported that students remembered and accurately reported their GPAs (correlation was .96 for 100 randomly selected students). To determine if the students in the freshman English classes were accurately reporting their GPAs, two students were randomly selected from each class. These 62 GPAs were checked with the official records. Official GPAs were then correlated with reported GPAs. Also an F-test and a t-test were used to determine if there were significant differences in population variances or means at the .10 level of significance.

An attempt was made by the investigator to increase the reliability of the scoring of the final examinations. Each instructor was given a list of recommendations for scoring essay examinations as proposed by Thorndike and Hagen (1969).

Two days before the final examination, each instructor was given a list of the names of the students in his class who completed the evaluation forms. Instructors wrote the final examination grades of their students on the list and returned it to the investigators by campus mail.

The final examination given to freshman English classes was an essay examination common to all classes. All students took the examination at the same time. Questions on the examination were chosen by a committee of three instructors of freshman English. Questions were submitted to this committee for their consideration by instructors teaching the freshman English class. Each instructor graded the papers from his class or classes. The instructors graded on a scale from F to A, using plus and minus signs. For the purpose of analyses, grades on the examination were converted to a 14-point scale (F=1 to A+=14).

One-way analyses of variance were used to determine if there were significant differences among the class means on the final examination. Separate analyses were used for the students responding to SIR and those responding to ISPI.

Three multivariate analyses of variance (MANOVAs) were used to determine if the vectors of the scores based on mean responses to the sections were significantly different among instructors. Since no computer program was available to handle 31 classes, 12 classes were randomly selected for the analyses. Hummel and Sligo (1971) reported empirical data that suggest that it is reasonable to follow up one-way MANOVAs with univariate analyses of variance (ANOVAs). To determine if there were significant differences among the instructors on each of the six rational sections of ISPI, the empirical sections of ISPI, and empirical sections of SIR for the 31 classes, ANOVAs were used. The level of significance for all MANOVAs and ANOVAs was set at .05.

Step-wise multiple correlational analyses were then carried out using the self-reported grade point averages (GPAs) together with mean responses to the empirical sections of the two instruments (obtained by the principal components) and to the rational sections of ISPI to predict mean end-of-course achievement scores. Since it seemed reasonable to assume that multiple correlation coefficients (Rs) based on class means would be more stable than Rs based on individual scores, .10 was set as the level of significance.

RESULTS OF DATA ANALYSES

After computing class means for each of the items on ISPI and SIR, a principal components with varimax rotation was used to determine the component loadings of the items. The six component solution to ISPI accounted for 80%

of the variance; however, one component contained only one item. Since one-item components are not generally stable, five components were then rotated. These five components accounted for 76% of the variance. Thirteen items had high loadings on two or more components and one item had loadings of less than .40 on each of the five components. These 14 items were eliminated from the further analyses.

The six component solution of SIR accounted for 73% of the variance. Three items were eliminated from further analyses since they had high loadings on more than one component

The varimax solution and the items which had significant loadings on each component are presented in Tables 1 and 2. Names assigned to the components of ISPI were: "Locus of Decision Making," "Course Relevance and Value," "Feelings of Acceptance or Tension," "Content Meaningfulness," and "Scope of Instructor Objectives." The six components of SIR were entitled: "Instructor's Planning and Learning Climate;" "Adjustment to Individual Student Needs;" "Work Load;" "Relevance of Methods and Materials;" "Texts, Tests and Student Interest;" and "Global Appraisals." These names are by no means definitive; they are only suggestive of the gist of the majority of items in that section. The components obtained using the class mean responses were used in the second phase of the study to determine criterion validity information.

GPAs were needed in the second phase of the study. To determine if the students were accurately reporting their GPAs, two students were randomly selected from each class, and their reported GPAs were checked against official records. The correlation of the official GPAs with reported GPAs was .94. This correlation was significant at the .01 level. The test for difference

in population means produced a t of .54, and the F-test for equality of the population variances was 1.16. Both of these results were non-significant at the .10 level. Results supported the hypotheses of no differences in the population means or the population variances of the two groups. Results, then supported the use of the self-reported GPAs as a predictor variable.

One-way analyses of variance were used to determine if there were significant differences among the class means on the final examination grades reported by the instructors. The obtained F-value for students responding to ISPI was 1.96, and the obtained F-value for students responding to SIR was 2.30. Both F-values were significant at the .05 level. Therefore, use of class mean scores on the final examination was supported. Three MANOVAs were used to determine if the vector scores based on mean responses to the sections were significantly different among instructors. The data are based on twelve randomly selected classes from the 31 classes. The likelihood ratio test statistics, chi-squares, for the instruments were as follows: 119.4 for the rational sections of ISPI, 122.5 for the empirical sections of ISPI, and 122.4 for the empirical sections of SIR. All of these were significant at the .05 level. Since the twelve classes were randomly selected, it can be assumed that similar results would have been obtained using all 31 classes. Results supported the use of class mean responses for the sections as the predictors.

Since results of the MANOVAs were significant, ANOVAs were used as follow ups. All 31 classes were used in the anlyses. At the .05 level of significance, follow-up ANOVAs revealed significant differences among average ratings for instructors for all rational sections of ISPI, all empirical sections of ISPI, and four of the six empirical sections of SIR. The F-values for section

3 of SIR--"Work Load"--and of 1.38 for section 4--"Relevance of Methods and Materials"--were not significant.

The class mean responses to each of the rational sections of ISPI, the empirical sections of ISPI, and the empirical sections of SIR (from Tables 3, 4, and 5) and class mean self-reported GPAs were then used to predict the class mean final examination scores in step-wise multiple correlation analyses (Kerlinger and Pedhazur, 1973). The correlation matrices for class mean section ratings, self-reported GPAs, and the final examination scores are presented in Tables 6, 7, and 8. At the .05 level, significant correlations existed between all mean section ratings. No significant correlations at the .05 level existed between the mean final examination scores and the mean section ratings.

In the first step-wise regression analyses, self-reported GPAs and six ISPI rational section predictor variables were entered. The multiple-R obtained using the GPAs and the six sections to predict final examination performance was .37; neither this value nor any of the other multiple-Rs was significant at the .10 level. Only four of the six ISPI empirical predictor variables together with self-reported GPAs were entered in the second step-wise regression. The largest multiple-R obtained was .42. Again, there was no empirical section of ISPI or combination of sections with self-reported GPA that contributed significantly to mean final examination scores of the freshman English classes.

A summary of the step-wise regression analysis for the SIR empirical sections and self-reported GPAs to predict class mean examination scores is provided in Table 9. R^2s (squared multiple-Rs) are also listed. These in-

dicate the proportion of variance that can be predicted from the predictor variables. Self-reported GPA and all six empirical sections of SIR were entered as predictor variables. Results of F-values to Remove revealed that in this study mean self-reported GPAs and two empirical sections of SIR (labeled "Adjustment to Individual Needs" and "Work Load") were statistically significant predictors of class mean examination performance. The multiple correlation obtained using the self-reported GPAs and the sections as predictors was .62. The proportion of variance accounted for by using these three predictors was .39.

SUMMARY OF CONCLUSIONS

The following conclusions were drawn from the data analyses:

1. Large amounts of variance were accounted for by using class mean responses to items as the units of analyses for both ISPI and SIR. A six-component solution of SIR accounted for 73% of the variance, and a five-component solution to ISPI accounted for 76% of the variance.

2. At the .05 level of significance, results of MANOVAS (based on twelve randomly selected classes) revealed that the vector scores based on the mean responses to the sections were significantly different among English instructors. At the .05 level of significance, follow-up ANOVAs revealed significant differences among average ratings for instructors for all rational sections of ISPI, all empirical sections of ISPI, and four of the six empirical sections of SIR.

3. For a stratified random sample of 62 students (2 from each of the 31 classes), tests for differences in actual and self-reported GPA means and variances were non-significant (.10 significance level).

4. The results of multiple correlational analyses revealed that none of the mean empirical section scores or of the mean rational section scores of ISPI was a significant predictor of end-of-course achievement.

5. At the .10 level of significance, results of the multiple correlational analysis revealed a statistically significant multiple correlation of .62 when mean self-reported GPAs and two empirical sections of SIR (labeled "Adjustment to Individual Student Needs" and "Work Load") were used to predict class mean end-of-course achievement.

IMPLICATIONS

While it is reasonable to believe that similar results would be obtained if the study were replicated in classes comparable to the freshman English classes of this study, generalizability to such classes and populations should be made with caution. Conceivably, different results could be obtained using different types of courses, different types of achievement examinations, and principal component scores rather than scores based on principal components. Taking these precautions into account, the results seem to have certain implications for the use, interpretation, and development of student-evaluation-of-college-instruction instruments. These implications are stated below:

1. Results of this study suggest that student appraisal of college instruction using SIR in conjunction with self-reported GPAs are of some help in predicting end-or-course final examination scores. Also examined from a practical viewpoint, the results of this study suggest careful examination of whatever instrument is used. If the effectiveness of an instructor is measured in terms of end-of-course achievement of his class, then college administrators should proceed with caution in using student-evaluation-of-college instruction

instruments to gauge instructor effectiveness.

2. Moreover, results of this study lend some support to the use of instruments developed empirically over those developed rationally. While all empirical sections of ISPI, all rational sections of ISPI, and four of the six empirical sections of SIR revealed significant differences in the ratings of the instructors on the sections, only two sections of SIR, together with self-reported GPAs bore a significant relationship to end-of-course final examination scores.

DISCUSSION

There are certain problems inherent in the design of this study that may have influenced this lack of relationship between student ratings and final examination scores. One problem involved the lack of anonymity of the ratings. While writers have refered to the necessity of anonymous ratings, it was necessary to have the students in this study identify themselves. This lack of anonymity may have affected the results; students may have responded differently if they had not had to supply their IDs.

Another factor that may have influenced the lack of relationship between ratings and achievement was the use of scores based on principal components rather than principal component scores. One reason for using scores based on principal components was that these scores are scores that instructors are familiar with. However, these section scores turned out to be highly inter-correlated but not significantly correlated with class mean achievement. Tables 6, 7, and 8 also indicate that the sections of SIR were not as highly inter-correlated as the rational and empirical sections of ISPI. Of the 15 inter-section correlations of SIR, five were less than .50 and only one was greater

than .80. With respect to ISPI rational sections, of the 15 inter-section correlations, nine were .80 or above and none were less than .60. Of the 10 inter-section correlations of ISPI empirical sections, two were .80 or above and only one was less than .60. These results could explain why the sections of SIR were better predictors of mean examination scores than the sections of ISPI. While principal component scores would have yielded components that were basically uncorrelated, the question remains as to whether they would have been significantly correlated with end-of-course achievement.

Another problem concerns the random halves of the English classes. The lack of relationship between ISPI sections and final examination performance could be attributed to the low correlation between the class mean self-reported GPA and the class mean score on the final examination. This correlation was only .09 (see Tables 7 and 8); whereas the same correlation for the random half of students responding to SIR was .36 (see Table 9). Although halves of classes were randomly chosen, these results indicate that the samples were not similar with respect to GPAs and final examination performance.

As has been referred to previously, one possible problem involved the essay examination. It may have been that the grades on the examination did not truly reflect achievement in the course.

REFERENCES

Centra, J. A. The student instructional report: Its development and uses. SIR report number 1. Princeton, New Jersey: Educational Testing Service, 1972.

Costin, F., Greenough, W. T., and Menges, R. J. Student ratings of college teaching: Reliability, validity, and usefulness. Review of Educational Research, 1971, 41, 511-535.

Eble, K. E. The Recognition and Evaluation of Teaching. Washington, D. C.: American Association of University Professors, One Dupont Circle, 1971.

Hummel, T. J. and Sligo, J. R. Empirical comparison of univariate and multivariate analysis of variance procedures. Psychological Bulletin, 1971, 76, 49-57.

Kerlinger, F. N., and Pedhazur, E. J. Multiple Regression in Behavioral Research. New York: Holt, Rinehart, and Winston, 1973, 290-295.

Miller, R. I. Evaluating Faculty Performance. San Francisco: Jossey-Bass, Inc., 1972.

Rayder, N. F. College student ratings of instructors. The Journal of Experimental Education, 1968, 37 (2), 76-81.

Scott, O. The Inventory of Student Perceptions of Instruction: Its history, characteristics and uses. Paper presented at the Twelfth Southeastern Invitational Conference on Measurement in Education, Blacksbury, Virginia, 1973.(a)

Scott, O. The measurement and use of student perceptions to improve college instruction. Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, Louisana, 1973.(b)

Thorndike, R. L. and Hagen, E. Measurement and Evaluation in Psychology and Education Third Edition. New York: John Wiley and Sons, Inc., 1969, 85-87, 166-169.

Tables 1-9, Inclusive

To Accompany

"A Comparison of the Criterion Validity of Two Types of Student Response Inventories for Appraising Instruction"*

Sidney E. Benton

North Georgia College

Owen Scott

The University of Georgia

*Paper presented on April 20, 1976, at the annual meeting of the National Council on Measurement in Education in San Francisco, California.

Table 1

Components and Component Loadings (Varimax Rotation) for ISPI

Using Class Mean Responses to Individual Items

(n=30 Classes)

		Factor Loading
Factor I: Locus of Decision Making		
23	The class has a "voice" in deciding what we do.	.87
18	In the class we try to understand points of view that differ from our own.	.85
27	The instructor tries to get more students to participate in making decisions which affect the class.	.85
28	The class has a "voice" in deciding what sequence we take up ideas, topics, and problems.	.81
25	The class has no "voice" in setting up acceptable standards for academic performance.	-.78
26	When problems arise, the class has a "voice" in working out solutions.	.78
24	The class has a "voice" in deciding how we do what is done.	.78
52	Instead of expecting every student to do the same thing, the instructor provides different activities for different students.	.73
63	Our class and the instructor look carefully at what we are learning and decide whether it is worth the time and effort we are spending on it.	.73
47	The class, with the instructor's help, selects a problem or area of interest to work on and then breaks it down to find out just what we want to learn and how to go about it.	.71
10	In this class we learn to express our ideas in ways which won't hurt other people's feelings or make them angry.	.70

Table 1 (Continued)

48	We identify and work on our problems in communicating--with other students and with the instructor.	.70
65	My grade depends primarily upon my improvement over my past performance.	.69
11	Through the way we interact in class we apply democratic values.	.68
17	In this class each student is accepted on his own merits.	.66
60	My grade is influenced by what is best for me as a person as well as by how much I have learned.	.64
20	What we do in class helps me understand myself better.	.63
29	The instructor gets help from resource persons when they can contribute to what the class is doing.	.63
12	In expressing ideas we learn to control our feelings.	.61
15	Our instructor encourages us to express different opinions and differing points of view on the ideas we discuss.	.60
21	The instructor is interested in me as a person.	.56

Factor II: Course Relevance and Value

46	What we learn is directly related to the work I plan to do when I graduate.	-.89
1	What we study does not help me plan my career.	.88
41	I try hard in this class because, to me, what I am trying to learn is worthwhile.	-.87
40	The class considers that what we are learning is worth learning.	-.71
42	Our assignments are uninteresting and are of little or no value.	.67
8	What we learn in impractical and of no use outside of class.	.65

Table 1 (Continued)

38	Because it is interesting, I do additional work outside of class that is not required.	-.60
51	By the time I have finished an activity or block of work, the things I've learned fit together to form a pattern that makes sense to me.	-.57
5	The instructor plans activities which apply what we have learned to situations that I might encounter later on in my vocation.	-.55
	Factor III: Feelings of Acceptance or Tension	
13	The atmosphere in this class is unfriendly.	.86
19	The instructor tries to force us to accept his ideas and interpretations.	.72
66	This instructor is the kind I admire and respect.	-.71
9	The instructor would rather have me think through something than memorize it.	-.68
62	This instructor's grading is fair.	-.68
16	This class makes me nervous.	.65
58	The way in which the instructor organizes ideas and activities is confusing.	.61
44	Instruction is planned in terms of the textbook sequence of content.	.57
	Factor IV: Content Meaningfulness	
32	When appropriate, we construct or devise our own instructional materials.	-.74
39	The content of this course is too difficult.	.72
49	The instructor's lectures and oral presentations are "over my head."	.72
50	The instructor does not speak clearly.	.70

Table 1 (Continued)

54	In moving from one idea to another, the instructor makes the connection clear.	-.69
64	My grade depends primarily on how well I do compared with the rest of the class.	.68
55	Our instructor explains things very clearly.	-.68
53	At the beginning of a topic or project I understand clearly what I am supposed to go about learning.	-.65
37	We have opportunities to carry out original assignments and projects.	-.64
35	We participate in a variety of class activities in addition to talking and listening.	-.64
2	In this class I developed skills and knowledge directly related to my plans after I finish college.	-.61
	Factor V: Scope of Instruction Objectives	
6	We learn to listen carefully to what other people say and to separate statements of fact from statements of opinion or attitude.	.75
4	We learn to express our ideas more precisely.	.71
57	At the end of an activity, unit or assignment, we summarize what we have learned.	.65
7	We learn such skills as identifying assumptions, reasoning logically from assumptions and testing conclusions.	.62
69	The instructor has a thorough knowledge of the subject matter.	.62
14	Class activities are planned so that every student who wishes to may make a contribution.	.60

Table 2

Components and Component Loadings for SIR

Using Class Mean Responses to Individual Items

(n=30 Classes)

		Factor Loading
Factor I: Instructor's Planning and Learning Climate		
3	The instructor used class time well.	.89
20	In my opinion, the instructor has accomplished (is accomplishing) his objectives.	.86
10	The instructor raised challenging questions or problems for discussion.	.83
1	The instructor's objectives for the course have been made clear.	.81
38	I would rate the overall value of class discussions...	.75
8	The instructor seemed genuinely concerned with students' progress and was actively helpful.	.70
7	The instructor encouraged students to think for themselves.	.69
14	The instructor summarized or emphasized major points in lectures or discussions.	.69
39	Compared to other instructors you have had (secondary school and college), how effective has the instructor been in this course:	.69
12	The instructor was well-prepared for each class.	.66
11	In this class I felt free to ask questions or express my opinions.	.61
34	Overall, I would rate the quality of the exams...	.61
Factor II: Adjustment to Individual Student Needs		
23	For me, the pace at which the instructor covered the material during the term was...	.84

Table 2 (Continued)

16	The scope of the course has been too limited; not enough material has been covered.	-.80
21	For my preparation and ability, the level of difficulty of this course was...	.68
19	The instructor was open to other viewpoints.	-.63
	Factor III: Work Load	
5	The instructor seemed to know when students didn't understand the material.	.72
18	I have been putting a good deal of effort into this course.	-.71
22	The work load for this course in relation to other courses of equal credit was...	-.67
	Factor IV: Relevance of Methods and Materials	
17	Examinations reflected the important aspects of the course.	.78
6	Lectures were too repetitive of what was in the textbook(s).	.70
9	The instructor made helpful comments on papers or exams.	.70
	Factor V: Texts, Tests and Student Interest	
32	Overall, I would rate the textbook(s)...	.79
15	My interest in the subject area has been stimulated by this course.	.72
13	The instructor told students how they would be evaluated in the course.	.55

Table 2 (Continued)

Factor VI: Global Appraisals		
35	I would rate the general quality of the lectures...	.85
36	I would rate the overall value of class discussions...	.64
24	To what extent did the instructor use examples or illustrations to help clarify the material?	.62

Table 3

Class Mean Self-Reported GPAs, Final Examination Scores, and Rational ISPI Section Scores

(n=31 Classes in Freshman English)

Class	n[a]	Mean Self-Reported GPA	Mean Final Exam	Section					
				I	II	III	IV	V	VI
1	10	2.52	8.70	3.42	3.57	3.28	3.53	3.44	3.13
2	8	2.72	7.56	3.55	3.86	3.23	3.59	3.70	3.45
3	8	2.89	8.25	3.53	3.70	3.02	3.43	3.21	3.29
4	9	2.70	7.78	3.61	3.71	3.45	3.58	3.63	3.54
5	8	2.73	10.00	4.06	4.24	3.79	3.99	4.04	4.00
6	3	2.62	6.00	3.47	3.82	3.50	3.67	3.76	3.81
7	6	2.68	10.17	3.78	3.78	3.53	3.71	3.67	3.72
8	7	2.88	7.71	3.19	3.62	3.43	3.35	3.52	3.65
9	7	2.93	11.29	3.57	3.48	3.??	3.54	3.37	3.04
10	8	2.73	8.38	3.60	3.67	3.75	3.61	3.62	3.21
11	10	2.53	9.30	3.23	3.01	3.02	2.97	3.00	3.03
12	9	3.11	11.00	3.51	3.86	3.30	3.51	3.38	3.32
13	9	2.89	11.00	3.13	3.78	3.67	3.67	3.45	3.65
14	11	2.57	8.09	4.00	4.15	3.85	4.08	3.96	3.94
15	8	2.77	10.88	3.71	4.04	3.52	3.86	3.53	3.20
16	7	2.35	13.00	3.70	4.10	3.57	3.51	3.89	3.55
17	6	2.32	11.00	3.60	3.85	3.72	3.91	3.91	3.67
18	7	2.41	8.71	3.51	3.39	3.55	3.51	3.51	3.27
19	11	3.02	9.09	3.28	3.50	3.12	3.28	3.28	3.17
20	6	3.21	9.50	3.25	3.19	3.36	3.68	3.64	3.41
21	6	2.75	8.50	3.10	3.42	3.25	3.17	3.15	2.76
22	7	2.95	9.57	3.66	3.10	2.81	3.38	3.50	3.51
23	8	2.64	9.50	3.21	3.11	3.27	3.17	3.29	3.16
24	9	2.71	8.33	2.96	2.88	2.52	2.92	3.01	2.81

Table 3 (Continued)

Class	n[a]	Mean Self-Reported GPA	Mean Final Exam	Section I	II	III	IV	V	VI
25	6	2.48	8.00	3.62	3.83	3.67	3.70	3.77	3.50
26	8	2.52	10.13	3.35	3.97	3.75	3.58	3.68	3.59
27	6	2.55	7.50	3.08	3.13	2.84	3.17	3.14	3.05
28	8	3.04	9.88	3.31	3.28	3.00	3.13	3.32	3.02
29	9	2.68	9.89	2.89	3.12	2.93	3.09	3.09	2.84
30	10	3.06	10.00	3.34	3.30	2.87	3.08	3.26	3.09
31	7	2.56	8.00	3.06	3.34	2.72	3.10	3.27	3.08

I--Instructional Objectives
II--Human Relations
III--Use of Instructional Resources
IV--Student Motivation
V--Meaningfulness of Content
VI--Measurement and Evaluation

[a]Data are based on a random half of each class.

Table 4

Class Mean Self-Reported GPAs, Final Examination Scores, and Empirical ISPI Section Scores

(n=31 Classes in Freshman English)

Class	n[a]	Mean Self-Reported GPA	Mean Final Exam	Section I	Section II	Section III	Section IV	Section V
1	10	2.52	8.70	3.23	3.22	3.71	3.53	3.77
2	8	2.72	7.56	3.36	3.22	4.17	3.67	4.17
3	8	2.89	8.25	3.42	3.29	3.58	3.43	3.58
4	9	2.70	7.78	3.42	3.19	3.91	3.76	3.98
5	8	2.73	10.00	4.01	3.85	4.37	3.74	4.40
6	3	2.62	6.00	3.76	3.67	3.58	3.67	3.72
7	6	2.68	10.17	3.52	3.41	3.77	3.80	4.14
8	7	2.88	7.71	3.22	2.86	4.11	3.55	3.74
9	7	2.93	11.29	3.02	2.98	3.69	3.43	3.79
10	8	2.73	8.38	3.53	3.28	3.73	3.52	3.81
11	10	2.53	9.30	2.76	2.68	3.26	3.34	3.33
12	9	3.11	11.00	3.45	2.97	3.93	3.50	3.85
13	9	2.89	11.00	3.44	2.86	4.12	3.72	3.57
14	11	2.57	8.09	3.77	3.74	4.22	4.18	4.41
15	8	2.77	10.88	3.55	3.40	4.11	3.81	4.15
16	7	2.35	13.00	3.68	3.30	4.37	3.88	4.00
17	6	2.32	11.00	3.65	3.71	4.02	3.88	4.06
18	7	2.41	8.71	3.13	3.14	3.78	3.71	3.81
19	11	3.02	9.09	3.16	2.97	3.61	3.36	3.55
20	6	3.21	9.50	2.90	3.17	3.77	3.71	3.78
21	6	2.75	8.50	3.03	2.72	3.31	3.26	3.83
22	7	2.95	9.57	2.86	3.16	3.93	3.40	4.05
23	8	2.64	9.50	2.86	2.69	3.76	3.19	3.69
24	9	2.71	8.33	2.65	2.73	3.00	3.11	2.95

Table 4 (Continued)

Class	n[a]	Mean Self-Reported GPA	Mean Final Exam	Section				
				I	II	III	IV	V
25	6	2.48	8.00	3.46	3.56	4.27	3.70	4.06
26	8	2.52	10.13	3.68	2.97	4.03	3.73	3.94
27	6	2.55	7.50	2.95	2.74	3.46	3.11	3.47
28	8	3.04	9.88	2.99	2.31	3.50	3.22	3.71
29	9	2.68	9.89	2.83	2.41	3.36	3.24	3.42
30	10	3.06	10.00	2.88	2.84	3.82	3.32	3.70
31	7	2.56	8.00	2.89	2.51	3.89	3.22	3.69

I--Locus of Decision Making
II--Course Relevance and Value
III--Feelings of Acceptance or Tension
IV--Content Meaningfulness
V--Scope of Instructor Objectives

[a]Data are based on a random half of each class.

Table 5

Class Mean Self-Reported GPAs, Final Examination Scores, and Empirical SIR Section Scores

(n=31 Classes in Freshman English)

Class	n[a]	Mean Self-Reported GPA	Mean Final Exam	Section					
				I	II	III	IV	V	VI
1	10	2.72	9.25	3.58	3.65	3.23	3.03	3.37	3.57
2	9	2.93	8.72	3.58	3.48	3.33	3.33	3.41	4.15
3	9	2.38	7.33	3.29	3.44	3.15	3.13	3.24	3.30
4	11	2.44	7.09	3.47	3.36	3.18	3.15	3.03	3.82
5	10	2.34	7.50	3.67	3.68	3.50	3.40	3.27	3.67
6	5	3.22	10.00	3.49	3.20	2.80	3.13	3.33	3.73
7	9	3.10	8.11	3.40	3.45	3.20	3.22	3.07	3.26
8	13	2.69	8.62	3.40	3.47	2.95	3.01	3.06	3.72
9	9	2.78	9.89	3.37	3.19	3.15	3.04	3.24	3.59
10	10	2.73	9.40	3.57	3.30	3.23	3.05	3.29	4.03
11	9	2.65	9.78	3.14	3.47	3.00	2.71	2.78	3.37
12	9	2.50	10.00	3.38	3.25	3.18	2.78	3.11	3.30
13	9	2.82	9.78	3.22	3.33	3.19	2.94	2.89	3.70
14	10	2.27	8.00	3.67	3.48	3.10	3.28	3.45	4.04
15	8	2.56	12.13	3.47	3.60	3.04	2.98	3.06	3.92
16	7	2.80	12.14	3.83	3.65	3.17	3.48	3.38	4.24
17	9	2.79	12.33	3.56	3.47	2.65	3.33	3.37	3.85
18	6	2.64	9.50	3.59	3.42	3.25	2.94	3.72	4.17
19	11	2.83	9.73	3.38	3.27	3.15	3.00	3.21	3.59
20	9	2.76	10.33	3.58	3.64	3.33	3.43	3.48	3.81
21	7	2.76	6.14	2.95	2.43	2.71	2.67	2.86	2.91
22	8	2.55	7.75	3.55	3.59	3.40	3.58	3.33	3.79
23	8	2.92	11.50	3.26	3.33	3.21	3.35	3.17	3.54
24	8	2.90	9.00	2.88	3.00	2.67	2.71	2.92	2.37

Table 5 (Continued)

Class	n[a]	Mean Self-Reported GPA	Mean Final Exam	Section I	II	III	IV	V	VI
25	6	2.50	7.33	3.97	3.54	3.23	3.56	3.22	4.00
26	9	2.63	8.78	3.70	3.56	3.07	3.50	3.09	4.11
27	9	2.80	8.33	2.68	3.09	3.15	2.80	2.93	2.50
28	7	2.59	8.43	3.37	3.73	3.24	2.95	3.12	3.57
29	6	2.71	9.00	2.79	3.26	2.89	3.09	2.89	2.72
30	9	3.39	10.33	3.17	2.96	3.09	2.91	3.00	3.07
31	10	2.81	8.40	3.42	3.22	3.37	3.12	3.28	3.87

I--Instructor's Planning and Learning Climate
II--Adjustment to Individual Student Needs
III--Work Load
IV--Relevance of Methods and Materials
V--Texts, Tests and Student Interest
VI--Global Appraisals

[a]Data are based on a random half of each class.

Table 6

Correlation Matrix for Mean ISPI Rational Section Scores, Self-Reported GPAs, and Final Examination Scores for 31 Classes in Freshman English

Section	1	2	3	4	5	6	GPA	Final Exam
1	1.00	.73*	.63*	.80*	.79*	.69*	-.14	.18
2		1.00	.80*	.83*	.80*	.73*	-.22	.19
3			1.00	.85*	.81*	.72*	-.30	.17
4				1.00	.88*	.80*	-.16	.14
5					1.00	.87*	-.28	.11
6						1.00	-.16	.01
GPA							1.00	.09
Final Exam								1.00

* Significant at .05 level

Section 1--Instructional Objectives

Section 2--Human Relations

Section 3--Use of Instructional Resources

Section 4--Motivation

Section 5--Meaningfulness of Content

Section 6--Measurement and Evaluation

Table 7

Correlation Matrix for Mean ISPI Empirical Section Scores, Self-Reported GPAs, and Final Examination Scores for 31 Classes in Freshman English

Section	1	2	3	4	5	GPA	Final Exam
1	1.00	.81*	.67*	.80*	.70*	-.29	.10
2		1.00	.56*	.79*	.73*	-.24	-.02
3			1.00	.71*	.79*	-.14	.29
4				1.00	.75*	-.28	.19
5					1.00	-.15	.16
GPA						1.00	.09
Final Exam							1.00

* Significant at .05 level

Section 1--Locus of Decision Making

Section 2--Course Relevance and Value

Section 3--Feelings of Acceptance or Tension

Section 4--Content Meaningfulness

Section 5--Scope of Instructor Objectives

Table 8

Correlation Matrix for Mean SIR Empirical Section Scores, Self-Reported GPAs, and Final Examination Scores of 31 Classes in Freshman English

Section	1	2	3	4	5	6	GPA	Final Exam
1	1.00	.66*	.44*	.71*	.70*	.90*	-.29	.11
2		1.00	.52*	.61*	.45*	.61*	-.42*	.21
3			1.00	.42*	.41*	.43*	-.27	-.18
4				1.00	.53*	.61*	-.18	.04
5					1.00	.67*	-.14	.17
6						1.00	-.26	.22
GPA							1.00	.36*
Final Exam								1.00

* Significant at .05 level

Section 1--Instructor's Planning and Learning Climate

Section 2--Adjustment to Individual Student Needs

Section 3--Work Load

Section 4--Relevance of Methods and Materials

Section 5--Texts, Tests and Student Interest

Section 6--Global Appraisals

Table 9

Summary of Step-wise Regression Analysis

SIR Empirical Sections and Self-Reported GPAs Predicting

Mean Final Examination Scores in 31 Classes in Freshman English

Step	Predictor Variable Entered	Multiple R	R^2	F to Remove
1	GPA	.36	.13	4.41*
2	Section 2	.54	.29	6.40*
3	Section 3	.62	.39	4.05*
4	Section 6	.64	.41	1.21
5	Section 1	.67	.45	1.86
6	Section 4	.68	.47	.61
7	Section 5	.69	.48	.61

* Significant at .10 level

Section 1--Instructor's Planning and Learning Climate

Section 2--Adjustment to Individual Student Needs

Section 3--Work Load

Section 4--Relevance of Methods and Materials

Section 5--Texts, Tests and Student Interest

Section 6--Global Appraisals